

# Diversity in Parametric Families of Number Fields

Yuri Bilu, Florian Luca

**Abstract** Let  $X$  be a projective curve defined over  $\mathbb{Q}$  and  $t \in \mathbb{Q}(X)$  a non-constant rational function of degree  $v \geq 2$ . For every  $n \in \mathbb{Z}$  pick a point  $P_n \in X(\bar{\mathbb{Q}})$  such that  $t(P_n) = n$ . A result of Dvornicich and Zannier implies that, for large  $N$ , among the number fields  $\mathbb{Q}(P_1), \dots, \mathbb{Q}(P_N)$  there are at least  $cN/\log N$  distinct; here  $c > 0$  depends only on the degree  $v$  and the genus  $\mathbf{g} = \mathbf{g}(X)$ . We prove that there are at least  $N/(\log N)^{1-\eta}$  distinct fields, where  $\eta > 0$  depends only on  $v$  and  $\mathbf{g}$ .

## 1 Introduction

*Everywhere in this paper “curve” means “smooth geometrically irreducible projective algebraic curve”.*

Let  $X$  be a curve over  $\mathbb{Q}$  of genus  $\mathbf{g}$  and  $t \in \mathbb{Q}(X)$  a non-constant rational function of degree  $v \geq 2$ . We fix, once and for all, an algebraic closure  $\bar{\mathbb{Q}}$ . Our starting point is the celebrated Hilbert Irreducibility Theorem.

**Theorem 1.1 (Hilbert).** *In the above set-up, for infinitely many  $n \in \mathbb{Z}$  the fiber  $t^{-1}(n) \subset X(\bar{\mathbb{Q}})$  is  $\mathbb{Q}$ -irreducible; that is, the Galois group  $G_{\bar{\mathbb{Q}}/\mathbb{Q}}$  acts on  $t^{-1}(n)$  transitively.*

This can also be re-phrased as follows: for every  $n \in \mathbb{Z}$  pick  $P_n \in t^{-1}(n)$ ; then for infinitely many  $n \in \mathbb{Z}$  we have  $[\mathbb{Q}(P_n) : \mathbb{Q}] = v$ .

---

Yuri Bilu

Institut de Mathématiques de Bordeaux, Université de Bordeaux & CNRS; e-mail: yuri@math.u-bordeaux.fr

Florian Luca

School of Mathematics, Wits University, Johannesburg and Centro de Ciencias Matematicas, UNAM, Morelia; e-mail: Florian.Luca@wits.ac.za

“Infinitely many” in the Hilbert Irreducibility Theorem means, in fact, “overwhelmingly many”: for sufficiently large positive  $N$  we have

$$|\{n \in [1, N] \cap \mathbb{Z} : t^{-1}(n) \text{ is reducible}\}| \leq c(v)N^{1/2}. \quad (1)$$

Everywhere in the introduction “sufficiently large” means “exceeding a certain positive number depending on  $X$  and  $t$ ”.

For the proof of (1) we invite the reader to consult Chapter 9 of Serre’s book [8]. See, in particular, Section 9.2 and the theorem on page 134 of [8], where (1) is proved with  $\mathbb{Q}$  replaced by an arbitrary number field and  $\mathbb{Z}$  by its ring of integers.

Hilbert’s Irreducibility Theorem, however, does not answer the following natural question: among the field  $\mathbb{Q}(P_n)$ , are there “many” distinct (in the fixed algebraic closure  $\bar{\mathbb{Q}}$ )? This question is addressed in the article of Dvornicich and Zannier [6], where the following theorem is proved (see [6, Theorem 2(a)]).

**Theorem 1.2 (Dvornicich, Zannier).** *In the above set-up, there exists a real number  $c = c(\mathbf{g}, v) > 0$  such that for sufficiently large integer  $N$  the number field  $\mathbb{Q}(P_1, \dots, P_N)$  is of degree at least  $e^{cN/\log N}$  over  $\mathbb{Q}$ .*

One may note that the statement holds true independently of the choice of the points  $P_n$ .

An immediate consequence is the following result.

**Corollary 1.3.** *In the above set-up, there exists a real number  $c = c(\mathbf{g}, v) > 0$  such that for every sufficiently large integer  $N$ , there are at least  $cN/\log N$  distinct fields among the number fields  $\mathbb{Q}(P_1), \dots, \mathbb{Q}(P_N)$ .*

Theorem 1.2 is best possible, as obvious examples show. Say, if  $X$  is (the projectivization of) the plane curve  $t = u^2$  and  $t$  is the coordinate function, then the field

$$\mathbb{Q}(P_1, \dots, P_N) = \mathbb{Q}(\sqrt{1}, \sqrt{2}, \dots, \sqrt{N}) = \mathbb{Q}(\sqrt{p} : p \leq N)$$

is of degree  $2^{\pi(N)} \leq e^{cN/\log N}$ .

On the contrary, Corollary 1.3 does not seem to be best possible. For instance, in the same example, if  $n$  runs the square-free numbers among  $1, \dots, N$  then the fields  $\mathbb{Q}(P_n) = \mathbb{Q}(\sqrt{n})$  are pairwise distinct. It is well-known that among  $1, \dots, N$  there are, asymptotically,  $\zeta(2)^{-1}N$  square-free numbers as  $N \rightarrow \infty$ .

We suggest the following conjecture.

**Conjecture 1.4 (Weak Diversity Conjecture).** Let  $X$  be a curve over  $\mathbb{Q}$  and  $t \in \mathbb{Q}(X)$  a non-constant  $\mathbb{Q}$ -rational function of degree at least 2. Then there exists a real number  $c > 0$  such that for every sufficiently large integer  $N$ , among the number fields  $\mathbb{Q}(P_1), \dots, \mathbb{Q}(P_N)$  there are at least  $cN$  distinct.

There is also a stronger conjecture, attributed in [6, 7] to Schinzel, which relates to Theorem 1.2 in the same way as Conjecture 1.4 relates to Corollary 1.3. To state it, we need to recall the notion of *critical value*.

We call  $\alpha \in \bar{\mathbb{Q}} \cup \{\infty\}$  a *critical value* (or a *branch point*) of  $t$  if the rational function<sup>1</sup>  $t - \alpha$  has at least one multiple zero in  $X(\bar{\mathbb{Q}})$ . It is well-known that any rational function  $t \in \bar{\mathbb{Q}}(X)$  has at most finitely many critical values, and that  $t$  has at least 2 distinct critical values if it is of degree  $v \geq 2$  (a consequence of the Riemann-Hurwitz formula). In particular, in this case  $t$  admits at least one *finite* critical value.

**Conjecture 1.5 (Strong Diversity Conjecture (Schinzel)).** In the set-up of Conjecture 1.4, assume that either  $t$  has at least one finite critical value not belonging to  $\mathbb{Q}$ , or the field extension  $\bar{\mathbb{Q}}(X)/\bar{\mathbb{Q}}(t)$  is not abelian. Then there exists a real number  $c > 0$  such that for every sufficiently large integer  $N$  the number field  $\mathbb{Q}(P_1, \dots, P_N)$  is of degree at least  $e^{cN}$  over  $\mathbb{Q}$ .

As Dvornicich and Zannier remark, the hypothesis in the Strong Diversity Conjecture is necessary. Indeed, when all critical values belong to  $\mathbb{Q}$  and the field extension  $\bar{\mathbb{Q}}(X)/\bar{\mathbb{Q}}(t)$  is abelian, it follows from Kummer's Theory that  $\mathbb{Q}(X)$  is contained in the field of the form  $L(t, (t - \alpha_1)^{1/e_1}, \dots, (t - \alpha_s)^{1/e_s})$ , where  $L$  is a number field,  $\alpha_1, \dots, \alpha_s$  are rational numbers and  $e_1, \dots, e_s$  are positive integers. Clearly, in this case the degree of the number field generated by  $P_1, \dots, P_N$  cannot exceed  $e^{cN/\log N}$  for some  $c > 0$ .

On the other hand, Conjecture 1.4 does hold [2] in the case excluded in Conjecture 1.5, when the finite critical values of  $t$  are all in  $\mathbb{Q}$ , and the field extension  $\bar{\mathbb{Q}}(X)/\bar{\mathbb{Q}}(t)$  is abelian. Hence, *the Strong Conjecture implies the Weak Conjecture*.

Dvornicich and Zannier [6, 7] obtain several results in favor of Schinzel's Conjecture. In particular, they show that Conjecture 1.5 holds true in the following cases:

- when  $t$  admits a critical value of degree 2 or 3 over  $\mathbb{Q}$ , see [6, Theorem 2(b)];
- when all finite critical values are in  $\mathbb{Q}$  and the Galois group of the normal closure of  $\bar{\mathbb{Q}}(X)$  over  $\bar{\mathbb{Q}}(t)$  is “sufficiently large” (for instance, symmetric or alternating), see [7].

A result of Corvaja and Zannier [3, Corollary 1] implies that, in the case when  $t$  has at least 3 zeros in  $X(\bar{\mathbb{Q}})$ , a number field  $K$  of degree  $v$  or less may appear as  $\mathbb{Q}(P_n)$  for at most  $c(X, t, v)$  possible  $n$ . In particular, the Weak Conjecture holds in this case (but the Strong Conjecture remains open).

We mention also the work of Zannier [9], who studies the following problem: given a number field  $K$ , how many fields among  $\mathbb{Q}(P_1), \dots, \mathbb{Q}(P_N)$  contain  $K$ ? He proves that, under suitable assumptions, the number of such fields is  $o(N^\varepsilon)$  as  $N \rightarrow \infty$  for any  $\varepsilon > 0$ .

In the present article we go a different way: instead of imposing additional restrictions on  $X$  and  $t$ , we work in full generality, improving on Corollary 1.3 quantitatively in the direction of Conjecture 1.4. Here is our principal result.

**Theorem 1.6.** *In the set-up of Conjecture 1.4, there exists a positive real number  $\eta = \eta(\mathbf{g}, v)$  such that for every sufficiently large integer  $N$ , among the number fields  $\mathbb{Q}(P_1), \dots, \mathbb{Q}(P_N)$  there are at least  $N/(\log N)^{1-\eta}$  distinct.*

The proof shows that  $\eta = 10^{-6}((\mathbf{g} + v) \log(\mathbf{g} + v))^{-1}$  would do.

<sup>1</sup> Here and everywhere below we use the standard convention  $t - \infty = t^{-1}$ .

## Plan of the article

In Section 2 we introduce the notation and recall basic facts, to be used throughout the article.

In Section 3 we review the argument of Dvornicich and Zannier, and explain how it should be modified for our purposes.

Sections 4, 5 and 6 are the technical heart of the article. In Section 4 and 5 we introduce a certain set of square-free numbers and study its properties. A key lemma used in Section 5 is proved in Section 6.

After all this preparatory work, the proof of Theorem 1.6 becomes quite transparent, see Section 7.

## Acknowledgments

During the work on this article Yuri Bilu was partially supported by the University of Xiamen, and by the binational research project MuDeRa, funded jointly by the French and the Austrian national science foundations ANR and FWF.

We thank Jean Gillibert and Felipe Voloch for useful discussions. We also thank the referees who carefully read the manuscript and detected several inaccuracies.

## 2 Notation and Conventions

Unless the contrary is stated explicitly, everywhere in the article:

- $n$  (with or without indexes) denotes a positive integer;
- $m$  (with or without indexes) denotes a square-free positive integer;
- $p$  (with or without indexes) denotes a prime number;
- $x, y, z$  denote positive real numbers.

We use the notation

$$p_{\max}(n) = \max\{p : p \mid n\}, \quad p_{\min}(n) = \min\{p : p \mid n\}.$$

As usual, we denote by  $\omega(n)$  (respectively  $\Omega(n)$ ) the number of prime divisors of  $n$  counted without (respectively, with) multiplicities.

For a separable polynomial  $F(T) \in \mathbb{Z}[T]$  we denote:

- $\Delta_F$  the discriminant of  $F$ ;
- $\mathcal{P}_F$  the set of  $p$  for which  $F(T)$  has a root mod  $p$ , and which do not divide  $\Delta_F$ .
- $\mathcal{M}_F$  the set of square-free integers composed of primes from  $\mathcal{P}_F$ .

By the Chebotarev Density Theorem, the set  $\mathcal{P}_F$  is of positive density among all the primes. We call it the *Chebotarev density* of  $F$  and denote it by  $\delta_F$ . Note that

$$\delta_F \geq \frac{1}{d}, \quad (2)$$

where  $d = \deg F$ .

### 3 The Argument of Dvornicich-Zannier

In this section we briefly review the beautiful ramification argument of Dvornicich and Zannier<sup>2</sup> and explain which changes are to be made therein to adapt it for proving Theorem 1.6.

Like in the introduction, in this section “sufficiently large” means “exceeding some quantity depending on  $X$  and  $t$ ”.

Let  $F(T) \in \mathbb{Z}[T]$  be the primitive separable polynomial whose roots are exactly the finite critical values of  $t$ , and let  $d = \deg F$ . Using the Riemann-Hurwitz formula, one bounds the total number of critical values by  $2\mathbf{g} - 2 + 2v$ , where  $\mathbf{g} = \mathbf{g}(X)$  is the genus of the curve  $X$ . Hence

$$d \leq 2\mathbf{g} - 2 + 2v. \quad (3)$$

The basic properties of the polynomial  $F(T)$  are summarized below.

- A** For sufficiently large  $p$ , if  $p$  ramifies in  $\mathbb{Q}(P)$  for some  $P \in t^{-1}(n)$  then  $p \mid F(n)$ .
- B** For sufficiently large  $p$ , if  $p \parallel F(n)$  then  $p$  ramifies in  $\mathbb{Q}(P)$  for some  $P \in t^{-1}(n)$ .
- C** For all  $p$  not dividing the discriminant  $\Delta_F$  (which is non-zero because  $F$  is a separable polynomial) the following holds: if for some  $n$  we have  $p^2 \mid F(n)$  then  $p \parallel F(n+p)$ .
- D** For every  $p \in \mathcal{P}_F$  there exists  $n \leq 2p$  such that  $p \parallel F(n)$ .
- E** When  $n$  is sufficiently large,  $F(n)$  has at most  $d$  prime divisors  $p \geq n/4$ .

Here properties **A** and **B** are rather standard statements linking geometric and arithmetical ramification, see [1, Theorem 7.8].

Property **C** is very easy: write

$$F(n+p) \equiv F(n) + F'(n)p \pmod{p^2}.$$

If  $p^2$  divides both  $F(n)$  and  $F(n+p)$  then  $p \mid F'(n)$ , which means that  $p$  must divide the discriminant  $\Delta_F$ , a contradiction.

Property **D** follows from **C**, and property **E** is obvious: if there are  $d+1$  such primes, then  $(n/4)^{d+1} \leq |F(n)|$ , which is impossible for large  $n$ .

One may also note that our definition of the polynomial  $F(T)$  is relevant only for properties **A** and **B**; the other properties hold for any separable polynomial  $F(T) \in \mathbb{Z}[T]$ .

---

<sup>2</sup> In [6] they trace it back to the work of Davenport et al [4] from sixties.

Now we are ready to sketch the proof of Theorem 1.2. Denote by  $K_n$  the number field  $\mathbb{Q}(t^{-1}(n))$ , generated by all the points in the fiber of  $n$ , and by  $L_n$  the compositum of the fields  $K_1, \dots, K_n$ . Then  $K_n$  is a Galois extension of  $\mathbb{Q}$  containing  $\mathbb{Q}(P_n)$ , and  $L_n$  is a Galois extension of  $\mathbb{Q}$  containing  $\mathbb{Q}(P_1, \dots, P_n)$ .

We call  $p$  *primitive* for some  $n$  if  $p$  ramifies in  $K_n$ , but not in  $L_{n-1}$ . The observations above have the following two consequences.

- F** Every sufficiently large  $p \in \mathcal{P}_F$  is primitive for some  $n \leq 2p$ .
- G** Every sufficiently large  $n$  has at most  $d$  primitive  $p \in [n/4, n]$ .

Here **F** follows from **B** and **D**, and **G** follows from **A** and **E**.

For a given  $N$  let  $S_N$  be the set of  $n$  with the property

$$n \text{ has a primitive } p \in [N/4, N/2].$$

It follows from **F** that  $S_N \subset [1, N]$ , and from **G**, the Chebotarev Theorem and the Prime Number Theorem that, for sufficiently large  $N$

$$|S_N| \geq \frac{1}{d} |\mathcal{P}_F \cap [N/4, N/2]| \geq \frac{\delta_F}{5d} \frac{N}{\log N}.$$

Furthermore, let  $S'_N$  be the subset of  $S_N$  consisting of  $n$  such that the fiber  $t^{-1}(n)$  is irreducible. The quantitative Hilbert Irreducibility Theorem 1 implies that, for large  $N$  we have  $|S_N \setminus S'_N| \leq c(v)N^{1/2}$ , which means that, for large  $N$ ,

$$|S'_N| \geq \frac{\delta_F}{6d} \frac{N}{\log N}.$$

It is clear that if  $n$  admits a primitive  $p$  then  $K_n$  is not contained in  $L_{n-1}$ . If, in addition to this, the fiber  $t^{-1}(n)$  is irreducible, then  $\mathbb{Q}(P_n)$  is not contained in  $\mathbb{Q}(P_1, \dots, P_{n-1})$ , because in this case  $K_n$  is the Galois closure (over  $\mathbb{Q}$ ) of  $\mathbb{Q}(P_n)$ . It follows that

$$[\mathbb{Q}(P_1, \dots, P_N) : \mathbb{Q}] \geq 2^{|S'_N|},$$

which, in view of (2) and (3), proves Theorem 1.2.

The (already mentioned in the Introduction) example of the curve  $u = t^2$  suggests that we can make progress towards Conjecture 1.4 replacing prime numbers in the argument above by (suitably chosen) square-free numbers. This means that we have to obtain analogues of properties **F** and **G** above with primes replaced by square-free numbers.

Let  $m$  be a square-free integer, and  $n$  an arbitrary integer. We say that  $m \parallel n$  if  $m \mid n$  and  $\gcd(m, n/m) = 1$ .

A “square-free analogue” of **F** is relatively easy: one uses the following lemma, which generalizes property **C**.

**Lemma 3.1.** *Let  $m$  be a square free positive integer, coprime with  $\Delta_F$  and such that  $p_{\min}(m) > \omega(m)$ . Assume that for some  $n$  we have  $m \mid F(n)$ . Then there exists  $\ell \in \{0, 1, \dots, \omega(m)\}$  such that  $m \parallel F(n + \ell m)$ .*

*Proof.* Assume the contrary: for every  $\ell \in \{0, 1, \dots, \omega(m)\}$  there exists  $p \mid m$  such that  $p^2 \mid f(n + \ell m)$ . By the box principle some  $p$  would occur for two distinct values  $\ell_1$  and  $\ell_2$ ; we will assume that  $0 \leq \ell_1 < \ell_2 \leq \omega(m)$ . We obtain

$$\begin{aligned} 0 &\equiv F(n + \ell_2 m) && \text{mod } p^2 \\ &\equiv F(n + \ell_1 m) + F'(n + \ell_1 m)(\ell_2 - \ell_1)m && \text{mod } p^2 \\ &\equiv F'(n + \ell_1 m)(\ell_2 - \ell_1)m && \text{mod } p^2. \end{aligned}$$

We have  $p \parallel m$  and, since

$$0 < \ell_2 - \ell_1 \leq \omega(m) < p_{\min}(m) \leq p,$$

we have  $p \nmid (\ell_2 - \ell_1)$ . Hence  $p \mid F'(n + \ell_1 m)$ , which implies that  $p \mid \Delta_F$ , a contradiction.  $\square$

Recall that the set  $\mathcal{P}_F$  consists of primes  $p$  not dividing the discriminant  $\Delta_F$  and such that  $F$  has a root mod  $p$ , and that  $\mathcal{M}_F$  is the set of square-free numbers composed of primes from  $\mathcal{P}_F$ . The following consequence is immediate.

**Corollary 3.2.** *Let  $m \in \mathcal{M}_F$  have the property  $p_{\min}(m) > \omega(m)$ . Then there exists  $n \leq m(\omega(m) + 1)$  such that  $m \parallel f(n)$ .*

*Proof.* The Chinese Remainder Theorem implies that for any  $m \in \mathcal{M}_F$  there exists  $n \leq m$  such that  $m \mid F(n)$ . Now use Lemma 3.1.  $\square$

Call  $m \in \mathcal{M}_F$  *primitive* for  $n$  if every  $p \mid m$  ramifies in  $K_n$ , and for every  $n' < n$  some  $p \mid m$  does not ramify in  $K_{n'}$ . Combining Corollary 3.2 with property **A**, we obtain a quite satisfactory generalization of property **F** to square-free numbers.

**Corollary 3.3.** *Let  $m$  be like in Corollary 3.2. Then  $m$  is primitive for some  $n \leq m(\omega(m) + 1)$ .*

Another task to accomplish is extending to square-free numbers property **G**. This is much more intricate, see Sections 4, 5 and 6.

## 4 A Special Set of Square-Free Numbers

In this section we fix a separable polynomial  $F(T) \in \mathbb{Z}[T]$  of degree  $d$  and a real number  $\varepsilon$  satisfying  $0 < \varepsilon \leq 1/2$ . “Sufficiently large” will always mean “exceeding a certain quantity depending on  $F$  and  $\varepsilon$ ”, and the constants implied by the “ $O(\cdot)$ ” and “ $\ll$ ” notation depend on  $F$  and  $\varepsilon$  unless the contrary is stated explicitly.

Recall that  $\mathcal{P}_F$  denotes the set of primes  $p$  not dividing the discriminant  $\Delta_F$  and such that  $F$  has a root mod  $p$ , and  $\mathcal{M}_F$  denotes the set of the square-free numbers

composed of primes from  $\mathcal{P}_F$ . Recall also that we denote by  $\delta = \delta_F$  the density of  $\mathcal{P}_F$ . We have, as  $x \rightarrow \infty$ ,

$$|\mathcal{P}_F \cap [0, x]| \sim \delta \frac{x}{\log x}, \quad |\mathcal{M}_F \cap [0, x]| \sim \gamma \frac{x}{(\log x)^{1-\delta}}$$

where  $\gamma = \gamma(F)$  is a certain positive real number.

Recall that, unless the contrary is stated explicitly, the letter  $n$  always denotes a positive integer,  $m$  a square-free positive integer and  $p$  a prime number.

We fix a big positive real number  $x$  and set

$$\kappa = \log \log x, \quad k = \lfloor \varepsilon \delta \log \log x \rfloor + 1, \quad y = e^{(\log x)^{1-\varepsilon}}.$$

Furthermore, we denote by  $\mathcal{M}_F(x)$  the set of  $m \in \mathcal{M}_F$  satisfying

$$\frac{x}{2\kappa} \leq m \leq \frac{x}{\kappa}, \quad p_{\max}(m) \geq x^{9/10}, \quad p_{\min}(m) \geq y, \quad \omega(m) = k + 1.$$

**Proposition 4.1.** *We have  $|\mathcal{M}_F(x)| = x(\log x)^{-1+\varepsilon\delta+o(1)}$  as  $x \rightarrow \infty$ .*

*Proof.* If  $m \in \mathcal{M}_F(x)$ , then  $m = Pm_1$ , where  $P = p_{\max}(m) \geq x^{9/10}$ . We denote by  $\mathcal{M}'_F(x)$  be the set of such  $m_1$ 's. Then  $\mathcal{M}'_F(x) \subset \mathcal{M}_F$  and for every  $m_1 \in \mathcal{M}'_F(x)$  we have

$$m_1 \leq x^{1/10}, \quad p_{\min}(m_1) \geq y, \quad \omega(m_1) = k. \quad (4)$$

Let us count suitable  $P$  for a fixed  $m_1$ . These are exactly the primes  $P \in \mathcal{P}_F$  from the interval  $[x/(2\kappa m_1), x/(\kappa m_1)]$  satisfying  $P \geq x^{9/10}$ . The following observations are crucial.

- Since  $m_1 \leq x^{1/10}$ , we have  $x/(\kappa m_1) > x^{4/5}$  for sufficiently large  $x$ . Hence, for a fixed  $m_1$ , the number of suitable  $P$  is bounded from above by

$$\pi\left(\frac{x}{\kappa m_1}\right) \ll \frac{x}{\kappa m_1 \log x}.$$

- If  $m_1 \leq x^{1/10}/2\kappa$  then every prime  $P \in \mathcal{P}_F \cap [x/(2\kappa m_1), x/(\kappa m_1)]$  is suitable. Hence, for a fixed  $m_1 \leq x^{1/10}/2\kappa$ , the number of suitable  $P$  is bounded from below by

$$\pi_F\left(\frac{x}{\kappa m_1}\right) - \pi_F\left(\frac{x}{2\kappa m_1}\right) = \left(\frac{\delta}{2} + o(1)\right) \frac{x}{\kappa m_1 \log(x/(\kappa m_1))} \gg \frac{x}{\kappa m_1 \log x}.$$

Here,  $\pi_F(T)$  counts the number of primes in  $\mathcal{P}_F \cap [0, T]$ .

Summing up over  $m_1 \in \mathcal{M}'_F(x)$ , we obtain

$$\frac{x}{\kappa \log x} \sum_{\substack{m_1 \in \mathcal{M}'_F(x) \\ m_1 \leq x^{1/10}/2\kappa}} \frac{1}{m_1} \ll |\mathcal{M}_F(x)| \ll \frac{x}{\kappa \log x} \sum_{m_1 \in \mathcal{M}'_F(x)} \frac{1}{m_1}. \quad (5)$$



We will show that the the right-hand side of (5) is bounded by  $x(\log x)^{-1+\varepsilon\delta+o(1)}$  from above, and the left-hand side from below.

The **upper bound** is easy:

$$\begin{aligned}
 \sum_{m_1 \in \mathcal{M}_F'(x)} \frac{1}{m_1} &\leq \frac{1}{k!} \left( \sum_{\substack{y \leq p \leq x \\ p \in \mathcal{P}_F}} \frac{1}{p} \right)^k \\
 &\ll \frac{1}{(k/e)^k} ((\delta + o(1)) \log \log x - (\delta + o(1)) \log \log y)^k \\
 &\ll \left( \frac{(e + o(1)) \varepsilon \delta \log \log x}{k} \right)^k \\
 &= (\log x)^{\varepsilon \delta + o(1)}
 \end{aligned} \tag{6}$$

as  $x \rightarrow \infty$ . Hence,  $|\mathcal{M}_F(x)| \leq x(\log x)^{-1+\varepsilon\delta+o(1)}$  as  $x \rightarrow \infty$ .

For the **lower bound**, set  $z = x^{(1/11 \log \log x)}$  and  $\mathcal{J} = [y, z]$  and consider the following two sets:

- the set  $\mathcal{M}_F''(x)$  of square-free numbers  $m_1$  with prime divisors in  $\mathcal{P}_F \cap \mathcal{J}$  and with  $\omega(m_1) = k$ ;
- the set  $\mathcal{N}_F''(x)$  of *non-square-free* numbers  $n_1$  with prime divisors in  $\mathcal{P}_F \cap \mathcal{J}$  and with  $\Omega(n_1) = k$ .

Clearly, every  $m_1 \in \mathcal{M}_F''(x)$  satisfies

$$m_1 \leq x^{k/(11 \log \log x)} < x^{1/11} \leq \frac{x^{1/10}}{2\kappa}$$

for large  $x$ . Hence the sum in the left-hand side of (5) can be bounded as follows:

$$\begin{aligned}
 \sum_{\substack{m_1 \in \mathcal{M}_F'(x) \\ m_1 \leq x^{1/10}/2\kappa}} \frac{1}{m_1} &\geq \sum_{m_1 \in \mathcal{M}_F''(x)} \frac{1}{m_1} \\
 &\geq \frac{1}{k!} \left( \sum_{p \in \mathcal{P}_F \cap [y, z]} \frac{1}{p} \right)^k - \sum_{n_1 \in \mathcal{N}_F''(x)} \frac{1}{n_1}.
 \end{aligned} \tag{7}$$

We need to estimate the first sum in (7) from below and the second sum from above.

For the first sum we use the same argument as before and get

$$\begin{aligned}
\frac{1}{k!} \left( \sum_{p \in \mathcal{P}_F \cap [y, z]} \frac{1}{p} \right)^k &\gg \frac{1}{\sqrt{k}} \frac{1}{(k/e)^k} ((\delta + o(1)) \log \log z - (\delta + o(1)) \log \log y)^k \\
&\gg \left( \frac{(e + o(1)) \varepsilon \delta \log \log x}{k} \right)^k \\
&= (\log x)^{\varepsilon \delta + o(1)}.
\end{aligned}$$

Now let us estimate the second sum in (7). Note that every  $n_1 \in \mathcal{N}_F''(x)$  satisfies  $n_1 \leq z^k < x$  and is divisible by the square of a prime  $p \geq y$ . Hence,  $n_1 = p^2 n_2$  for some  $n_2 \leq x$ . It follows that

$$\sum_{n_1 \in \mathcal{N}_F''(x)} \frac{1}{n_1} \leq \left( \sum_{p \geq y} \frac{1}{p^2} \right) \left( \sum_{n_2 \leq x} \frac{1}{n_2} \right) \ll \frac{\log x}{y} = o(1)$$

as  $x \rightarrow \infty$ .

Putting all the estimates together, we conclude that

$$|\mathcal{M}_F(x)| \gg \frac{x(\log x)^{\varepsilon \delta + o(1)}}{\log x \log \log x} = \frac{x}{(\log x)^{1 - \varepsilon \delta + o(1)}}$$

as  $x \rightarrow \infty$ , which is what we wanted.  $\square$

## 5 Greedy and Generous Square-free Numbers

We retain the notation and set-up of Section 4.

As we have already remarked in Section 3, the Chinese Remainder Theorem implies that for any  $m \in \mathcal{M}_F$  there exists a positive integer  $n$  such that  $m \mid F(n)$ . Moreover, if  $m \in \mathcal{M}_F(x)$  then we can choose such  $n$  satisfying  $n \leq x$ . Of course, there can be several  $n$  with this property; pick one of them and call it  $n_m$ .

Thus, for every  $m \in \mathcal{M}_F(x)$  we pick  $n_m \leq x$  such that  $m \mid f(n_m)$ ; we fix this choice of the numbers  $n_m$  until the end of this section.

It might happen that  $n_m = n_{m'}$  for distinct  $m, m' \in \mathcal{M}_F(x)$ . It turns out, however, that, with a suitable choice of our parameter  $\varepsilon$ , the repetitions are “not too frequent”.

Call  $m \in \mathcal{M}_F(x)$  *generous* if it shares its  $n_m$  with at least  $6d$  other elements of  $\mathcal{M}_F(x)$ , and *greedy* otherwise.

**Proposition 5.1.** *Specify*

$$\varepsilon = \frac{1}{10^3 \log(2d)}. \quad (8)$$

*Then for sufficiently large  $x$  at least half of the elements of the set  $\mathcal{M}_F(x)$  are greedy. In particular,*

$$|\{n_m : m \in \mathcal{M}_F(x)\}| \geq \frac{1}{12d} |\mathcal{M}_F(x)|.$$

The crucial tool in the proof of this proposition is the following lemma, which might be viewed as a partial “square-free” version of Property **E** from Section 3. We cannot affirm that  $F(n)$  has “few” divisors in  $\mathcal{M}_F$  for all  $n$ ; but we can affirm that, with “few” exceptions,  $F(n)$  has “few” divisors in  $\mathcal{M}_F(x)$ .

**Lemma 5.2.** *For sufficiently large  $x$ , the set of  $n \leq x$  such that  $F(n)$  has more than  $6d$  divisors in  $\mathcal{M}_F(x)$ , is of cardinality at most  $x(\log x)^{-2+30\varepsilon \log(2d)}$ .*

We postpone the proof of this lemma until Section 6.

### 5.1 Initializing the Proof of Proposition 5.1

Starting from this subsection we work on the proof of Proposition 5.1.

We set  $\mathcal{J} = [y, x]$  and we try to understand the function  $\omega_{\mathcal{J}}(F(n))$ , where  $\omega_{\mathcal{J}}(\cdot)$  is the number of prime factors of the argument in the interval  $\mathcal{J}$ . We split  $n$  into three sets as follows.

- (i)  $E(x)$  (enormous), which is the set of  $n \leq x$  for which

$$\omega_{\mathcal{J}}(F(n)) \geq 3d(\log \log x)^2.$$

- (ii)  $L(x)$  (large), which is the set of  $n \leq x$  for which

$$\omega_{\mathcal{J}}(F(n)) \in [10^5 d^2 \log \log x, 3d(\log \log x)^2].$$

- (iii)  $R(x)$  (reasonable), which is the set of  $n \leq x$  such that

$$\omega_{\mathcal{J}}(F(n)) \leq 10^5 d^2 \log \log x.$$

For the purpose of this argument, if  $s = \omega_{\mathcal{J}}(F(n))$  then we denote all the prime factors of  $F(n)$  in  $\mathcal{J}$  by  $p_1 < p_2 < \dots < p_s$ .

We will use the multiplicative function  $\rho_F$ , defined for a positive integer  $u$  by

$$\rho_F(u) = |\{0 \leq n \leq u-1 : F(n) \equiv 0 \pmod{u}\}|. \quad (9)$$

Clearly,  $\rho_F(m) \leq d^{\omega(m)}$  holds for all squarefree positive integers  $m$ .

### 5.2 Counting $m$ with $n_m \in E(x)$

Since  $|F(n)| \ll n^d \ll x^d$  it follows that in case (i), if we put  $U = \lfloor (\log \log x)^2 \rfloor$ , then  $p_1 \cdots p_U \leq x^{1/2}$  for large  $x$ .

To count  $E(x)$ , fix  $p_1 < p_2 < \dots < p_U$  all in  $\mathcal{J}$  and let us count the number of  $n \leq x$  such that  $m_1 \mid f(n)$ , where  $m_1 = p_1 \cdots p_U$ . The number of such  $n$  is

$$\frac{\rho_F(m_1)}{m_1}x + O(\rho_F(m_1)) \ll \frac{d^{\omega(m_1)}}{m_1}x + d^{\omega(m_1)} \ll \frac{d^{\omega(m_1)}}{m_1}x. \quad (10)$$

In the middle of (10), the first term  $d^{\omega(m_1)}x/m_1$  dominates because  $m_1 \leq x^{1/2}$ .

We sum up over the possible  $m_1$  getting

$$|E(x)| \ll xd^U \sum \frac{1}{m_1}, \quad (11)$$

where the sum runs over all square-free  $m_1$  satisfying  $\omega(m_1) = U$  and having all prime divisors in  $\mathcal{J}$ . We estimate this sum by the multinomial coefficient trick, already used in the proof of Proposition 4.1:

$$\sum \frac{1}{m_1} \ll \frac{1}{U!} \left( \sum_{y \leq p \leq x} \frac{1}{p} \right)^U \ll \left( \frac{3 \log \log x}{U} \right)^U$$

This gives us the estimate

$$|E(x)| \ll x \left( \frac{3d \log \log x}{U} \right)^U,$$

which, with our definition  $U = \lfloor (\log \log x)^2 \rfloor$ , implies that

$$|E(x)| \leq xe^{-(1+o(1))(\log \log x)^2 \log \log \log x}$$

as  $x \rightarrow \infty$ .

Having bounded  $|E(x)|$ , we may now estimate the number of  $m$  such that  $n_m \in E(x)$ . For each  $n \leq x$  we have  $|F(n)| \ll n^d \leq x^d$  which implies that, for large  $x$ , we have  $\omega_{\mathcal{J}}(F(n)) \leq \log x$ . Thus, for large  $x$ , the divisor  $m \mid F(n)$  with  $\omega(m) = k$  can be chosen in at most

$$\binom{\lfloor \log x \rfloor}{k+1} \leq (\log x)^{k+1} \ll e^{2(\log \log x)^2}$$

ways. This implies that, as  $x \rightarrow \infty$ ,

$$\begin{aligned} |\{m \in \mathcal{M}_F(x) : n_m \in E(x)\}| &\leq |E(x)| e^{2(\log \log x)^2} \\ &\leq xe^{-(1+o(1))(\log \log x)^2 \log \log \log x}. \end{aligned}$$

Proposition 4.1 implies that this is  $o(|\mathcal{M}_F(x)|)$  as  $x \rightarrow \infty$ .

### 5.3 Counting $m$ with $n_m \in L(x)$

Let us deal with (ii) now. We let  $i_0$  and  $i_1$  be the maximal and the minimal positive integers such that  $2^{i_0} \leq 10^5 d$  and  $2^{i_1} \geq 3(\log \log x)$ , respectively. Clearly,  $i_1 - i_0 = O(\log \log \log x)$ . Consider an integer  $j \in [i_0, i_1 - 1]$  and denote by  $L_j(x)$  the subset of  $L(x)$  consisting of  $n$  such that

$$\omega_{\mathcal{J}}(F(n)) \in [2^j d \log \log x, 2^{j+1} d \log \log x].$$

We revisit the previous argument. We now take  $U = \lfloor 2^{j-1} \log \log x \rfloor$ , and let  $m_1 = p_1 \cdots p_U$ . Then  $m_1^{2^d} \leq |F(n)| \ll x^d$ , therefore  $m_1 \ll x^{1/2}$ . Now exactly as before we prove that

$$|L_j(x)| \ll x \left( \frac{3d \log \log x}{U} \right)^U,$$

which, with our definition  $U = \lfloor 2^{j-1} \log \log x \rfloor$ , implies that

$$|L_j(x)| \ll \frac{x}{(\log x)^{2^{j-2} \log(2^{j-2}/3d)}}.$$

Since

$$\log \frac{2^{j-2}}{3d} \geq \log \frac{2^{i_0-2}}{3d} \geq \log \frac{10^5 d}{24d} \geq 8,$$

we have

$$|L_j(x)| \ll \frac{x}{(\log x)^{2^{j+1}}}.$$

On the other hand, for  $n \in L_j(x)$  we have  $\omega_{\mathcal{J}}(F(n)) \leq 2^{j+1} d \log \log x$ . It follows that, for large  $x$ , the number of choices for  $m$  for a given  $n \in L_j(x)$  is at most

$$\begin{aligned} \binom{\lfloor 2^{j+1} d \log \log x \rfloor}{k+1} &\leq \frac{(2^{j+1} d \log \log x)^{k+1}}{(k+1)!} \\ &\leq \left( \frac{2^{j+3} d}{\delta \varepsilon} \right)^{2\delta \varepsilon \log \log x} \\ &= (\log x)^{2\delta \varepsilon \log(2^{j+3} d / \delta \varepsilon)}. \end{aligned} \tag{12}$$

Since

$$\frac{2^{j-1}}{\delta \varepsilon} \geq 2^{i_0-1} \geq \frac{10^5 d}{4} \geq 10^4 d,$$

we have

$$\frac{2^{j-1}}{\delta \varepsilon} \geq 2 \log \frac{2^{j-1}}{\delta \varepsilon} \geq \log \left( \frac{2^{j-1}}{\delta \varepsilon} \cdot 10^4 d \right) \geq \log \frac{2^{j+3} d}{\delta \varepsilon},$$

which shows that the exponent in (12) does not exceed  $2^j$ .

Thus, for large  $x$

$$|\{m \in \mathcal{M}_F(x) : n_m \in L_j(x)\}| \leq |L_j(x)|(\log x)^{2j} \ll \frac{x}{(\log x)^{2j}} \leq \frac{x}{(\log x)^2},$$

because  $2^j \geq 2^{i_0} \geq 10^5 d/2 \geq 2$ . Since there are  $O(\log \log \log x)$  possible  $j$ , we conclude that

$$|\{m \in \mathcal{M}_F(x) : n_m \in L(x)\}| \ll \frac{x \log \log \log x}{(\log x)^2},$$

which is again  $o(|\mathcal{M}_F(x)|)$  as  $x \rightarrow \infty$ .

Thus, we have proved that

$$|\{m : n_m \in E(x) \cup L(x)\}| = o(|\mathcal{M}_F(x)|) \quad (13)$$

as  $x \rightarrow \infty$ .

### 5.4 Completing the proof

We are ready now to complete the proof of Proposition 5.1. It remains to deal with  $n \in R(x)$ . If  $n \in R(x)$ , then  $\omega_{\mathcal{J}}(F(n)) \leq 10^5 d^2 \log \log x$ . Thus, for fixed  $n \in R(x)$  we have

$$\begin{aligned} |\{m \in \mathcal{M}_F(x) : n_m = n\}| &\leq \binom{\lfloor 10^5 d^2 \log \log x \rfloor}{k+1} \\ &\leq \frac{(10^5 d^2 \log \log x)^{k+1}}{(k+1)!} \\ &\leq \left( \frac{10^6 d^2}{\varepsilon \delta} \right)^{2\varepsilon \delta \log \log x} \\ &= (\log x)^{2\varepsilon \delta \log(10^6 d^2 / \varepsilon \delta)}. \end{aligned} \quad (14)$$

Now we are done: Lemma 5.2 combined with estimate (14) implies that there exists at most

$$\frac{x}{(\log x)^{2-30\varepsilon \log(2d)-2\varepsilon \delta \log(10^6 d^2 / \varepsilon \delta)}} \quad (15)$$

generous  $m \in \mathcal{M}_F(x)$  with the property  $n_m \in R(x)$ . When  $\varepsilon$  is chosen as in (8), a quick calculation shows that

$$30\varepsilon \log(2d) + 2\varepsilon \delta \log\left(\frac{10^6 d^2}{\varepsilon \delta}\right) < \frac{1}{2}.$$

Hence (15) is  $o(|\mathcal{M}_F(x)|)$  as  $x \rightarrow \infty$ . In particular, when  $x$  is sufficiently large, at least half of elements of  $\mathcal{M}_F(x)$  are greedy.  $\square$

It remains to prove Lemma 5.2.

## 6 Proof of Lemma 5.2

We keep the notation of Section 4, especially  $y = \exp((\log x)^{1-\varepsilon})$ .

### 6.1 Two Simple Lemmas

Let  $A$  be the subset of  $\mathcal{M}_F$  consisting of  $m$  with  $p_{\min}(m) \geq y$ . We study the set  $A(z) = A \cap [y, z]$  for  $z \in [y, x]$ .

**Lemma 6.1.** *When  $x$  is sufficiently large we have  $|A(z)| \leq z(\log x)^{-1+3\varepsilon}$  for all  $z \in [y, x]$ .*

*Proof.* Let  $g(n)$  be the characteristic function of  $A$ . Then for any  $z > 1$  we have

$$\sum_{p \leq z} g(p) \log p \leq 2z,$$

and  $g(p^n) = 0$  for  $n \geq 2$ . Using Lemma 9.6 on page 138 in [5], we obtain

$$|A(z)| = \sum_{n \leq z} g(n) \leq 3 \frac{z}{\log z} \sum_{n \in A(z)} \frac{1}{n}. \quad (16)$$

Clearly,  $\log z \geq (\log x)^{1-\varepsilon}$  for  $z \in [y, x]$ . As for the sum above, we have

$$\sum_{n \in A(z)} \frac{1}{n} \leq \prod_{y \leq p \leq z} \left(1 + \frac{1}{p}\right) \leq (\log x)^{\varepsilon+o(1)}$$

as  $x \rightarrow \infty$ . Together with (16) this finishes the proof.  $\square$

**Lemma 6.2.** *Assuming  $x$  sufficiently large, for  $y \leq a \leq b \leq x$  we have*

$$\sum_{\substack{a \leq n \leq b \\ n \in A}} \frac{1}{n} \leq \frac{\log b - \log a + 1}{(\log x)^{1-3\varepsilon}}.$$

*Proof.* Using Abel summation and Lemma 6.1, we obtain

$$\begin{aligned}
\sum_{\substack{a \leq n \leq b \\ n \in A}} \frac{1}{n} &= \int_a^b \frac{d|A(z)|}{z} \\
&= \frac{|A(b)|}{b} - \frac{|A(a)|}{a} + \int_a^b \frac{|A(z)|}{z^2} dz \\
&\leq \frac{|A(b)|}{b} + \frac{1}{(\log x)^{1-3\varepsilon}} \int_a^b \frac{dz}{z} \\
&\leq \frac{1}{(\log x)^{1-3\varepsilon}} + \frac{\log b - \log a}{(\log x)^{1-3\varepsilon}},
\end{aligned}$$

as wanted.  $\square$

## 6.2 Cliques

Starting from this subsection we begin the proof of Lemma 5.2. Recall that every  $m \in \mathcal{M}_F(x)$  writes as  $m = m_1 P$ , where  $P = p_{\max}(m) \geq x^{9/10}$ . As in Section 4 we denote by  $\mathcal{M}'_F(x)$  the set of all  $m_1$  obtained this way. They satisfy (4), which will be used in the sequel without special reference.

Let  $n \leq x$  be such that  $F(n)$  has at least  $6d$  distinct divisors in  $\mathcal{M}_F(x)$ . Write each of them  $m_1 P$  as above and let  $s$  be the number of such  $P$ . Then  $x^{9s/10} \leq |f(n)| \ll x^d$ , so  $s \leq 10d/9 + o(1)$  as  $x \rightarrow \infty$ . In particular,  $s < 2d$  for large  $x$ . Hence among the  $6d$  divisors there are three with the same  $P$ ; write them  $m_1 P$ ,  $m_2 P$  and  $m_3 P$ .

Let us call an (unordered) triple of pairwise distinct  $m_1, m_2, m_3 \in \mathcal{M}'_F(x)$  a *clique* if there exists a prime  $P \geq x^{9/10}$  such that  $m_1 P, m_2 P, m_3 P \in \mathcal{M}_F(x)$ . If  $\{m_1, m_2, m_3\}$  is a clique then  $m_1 P, m_2 P, m_3 P \in [x/(2\kappa), x/\kappa]$ . This implies that in a clique we have

$$\frac{m_j}{2} \leq m_i \leq 2m_j \quad (17)$$

for any  $i, j$ . In addition to this, since  $m_1, m_2, m_3$  in a clique are square-free with the same number of prime factors, we have

$$\gcd(m_i, m_j) < m_i < [m_i, m_j], \quad (i \neq j). \quad (18)$$

where  $[\dots]$  denotes the least common multiple. We will repeatedly use these properties.

## 6.3 The Sum over Cliques

To prove the lemma, it suffices to estimate the number of  $n$  such that  $F(n)$  has three distinct divisors forming a clique. When a clique  $\{m_1, m_2, m_3\}$  is fixed, the number of such  $n$  is at most



$$\frac{\rho_F([m_1, m_2, m_3])}{[m_1, m_2, m_3]}x + O(\rho_F([m_1, m_2, m_3])), \quad (19)$$

where  $\rho_F(\cdot)$  is defined in (9). When  $x$  is large, we have

$$\omega([m_1, m_2, m_3]) \leq 3k \leq 4\varepsilon \log \log x,$$

which implies

$$\rho_F([m_1, m_2, m_3]) \leq d^{\omega([m_1, m_2, m_3])} \leq (\log x)^{4\varepsilon \log d}.$$

Further, since  $m_i \leq x^{1/10}$ , we have  $[m_1, m_2, m_3] \leq x^{3/10} \leq x^{1/2}$ . It follows that in (19) the first term dominates over the second one, and the number of our  $n$  (for the fixed  $m_1, m_2, m_3$ ) is bounded, for large  $x$ , by

$$x(\log x)^{5\varepsilon \log d} \frac{1}{[m_1, m_2, m_3]}.$$

Hence the total number of  $n$  (for all possible choices of  $m_1, m_2, m_3$ ) is bounded by  $x(\log x)^{5\varepsilon \log d} S$ , where

$$S = \sum_{\{m_1, m_2, m_3\}} \frac{1}{[m_1, m_2, m_3]},$$

the summation being over all cliques. The rest of the argument is estimating this sum  $S$ .

We write  $S = S' + S''$ , where  $S'$  is the sum over the cliques with the property

$$\text{there is a relabeling of the indices such that } [m_1, m_2] < [m_1, m_2, m_3], \quad (20)$$

and  $S''$  is over the cliques such that

$$[m_1, m_2] = [m_1, m_3] = [m_2, m_3] = [m_1, m_2, m_3]. \quad (21)$$

## 6.4 Estimating $S'$

We are starting now to estimate  $S'$ . All cliques appearing in this subsection satisfy (20).

### 6.4.1 The estimate with $m_1$ and $m_2$ fixed

Fix  $m_1$  and  $m_2$ . Then  $m_3 \nmid [m_1, m_2]$  by (20). Set  $u = \gcd(m_3, [m_1, m_2])$ . With  $m_1$  and  $m_2$  being fixed, there are at most

$$2^{2k} \ll (\log x)^{3\epsilon\delta}$$

choices for  $u$  as a divisor of  $[m_1, m_2]$ .

Writing  $m_3 = uv$ . Clearly,  $v \in A$ , where  $A$  is the set from Subsection 6.1. Using (17), we obtain  $m_1/(2u) \leq v \leq 2m_1/u$ . Since  $u$  is a proper divisor of  $m_3$ , we also have  $v > 1$ , which implies  $v \geq y$ , because  $v \in A$ . Also, clearly  $v \leq m_3 \leq x$ . This shows that

$$\max \left\{ y, \frac{m_1}{2u} \right\} \leq v \leq \min \left\{ x, 2\frac{m_1}{u} \right\}. \quad (22)$$

We have  $[m_1, m_2, m_3] = [m_1, m_2]v$ . Thus, assuming  $m_1$  and  $m_2$  fixed, and summing up over all possible  $m_3$ , we get

$$\begin{aligned} \sum \frac{1}{[m_1, m_2, m_3]} &\leq \frac{1}{[m_1, m_2]} \sum_{u|[m_1, m_2]} \sum_{v \in A \text{ satisfying (22)}} \frac{1}{v} \\ &\ll \frac{1}{[m_1, m_2](\log x)^{1-4\epsilon}} \sum_{u|[m_1, m_2]} 1 \\ &\ll \frac{1}{(\log x)^{1-8\epsilon} [m_1, m_2]}. \end{aligned} \quad (23)$$

Here, in the inner sum in (23), we applied Lemma 6.2 with the choices

$$b = \min \left\{ x, \frac{2m_1}{u} \right\}, \quad a = \max \left\{ y, \frac{m_1}{2u} \right\},$$

and we used the fact that  $\log b - \log a \ll 1$ .

#### 6.4.2 The estimate with $m_1$ fixed

We now fix  $m_1$  and vary  $m_2$ . This time we set  $u = \gcd(m_1, m_2)$  and again write  $m_2 = uv$ . There are at most  $2^k \ll (\log x)^{2\epsilon\delta}$  choices for  $u$ . Furthermore, it follows from (18) that  $u$  is a proper divisor of  $m_2$ , which implies  $v > 1$ . Thus, our  $v$  again belongs to the set  $A$  and satisfies (22).

Keeping  $m_1$  fixed, we argue as above:

$$\begin{aligned} \sum \frac{1}{[m_1, m_2]} &= \frac{1}{m_1} \sum_{u|m_1} \sum_{v \in A \text{ satisfying (22)}} \frac{1}{v} \\ &\ll \frac{1}{m_1(\log x)^{1-4\epsilon}} \sum_{u|m_1} 1 \\ &\ll \frac{1}{m_1(\log x)^{1-7\epsilon}}. \end{aligned}$$

### 6.4.3 Estimating $S'$

Now we are ready to estimate  $S'$ :

$$S' \ll \frac{1}{(\log x)^{2-15\epsilon}} \sum_{m_1 \in \mathcal{M}'_f} \frac{1}{m_1} \ll \frac{1}{(\log x)^{2-17\epsilon}},$$

where for the last estimate we used (6).

## 6.5 Estimating $S''$

Now let  $\{m_1, m_2, m_3\}$  be a clique satisfying (21). Setting  $u = \gcd(m_1, m_2, m_3)$  and  $v_i = [m_1, m_2, m_3]/m_i$ , we obtain

$$\begin{aligned} m_1 &= uv_2v_3, & m_2 &= uv_1v_3, & m_3 &= uv_1v_2, \\ [m_1, m_2] &= [m_1, m_3] = [m_2, m_3] = [m_1, m_2, m_3] &= uv_1v_2v_3. \end{aligned}$$

We again use (18) to obtain  $v_i > 1$ , which implies  $v_i \geq y$  because  $v_i \in A$ . Also,  $v_i \leq x$ . Together with (17) this gives

$$\max\left\{y, \frac{v_1}{2}\right\} \leq v_i \leq \min\{x, 2v_1\} \quad (i = 2, 3). \quad (24)$$

It follows that

$$S'' \leq \sum_{\substack{u, v_1, v_2, v_3 \in A \\ \text{satisfying (24)}}} \frac{1}{uv_1v_2v_3}.$$

When  $u$  and  $v_1$  are fixed, we have

$$\sum_{\substack{v_2, v_3 \in A \\ \text{satisfying (24)}}} \frac{1}{uv_1v_2v_3} \leq \frac{1}{uv_1} \left( \sum_{\substack{v \in A \\ \max\{y, v_1/2\} \leq v \leq \min\{x, 2v_1\}}} \frac{1}{v} \right)^2,$$

and the squared sum can be estimated, using Lemma 6.2, as  $O((\log x)^{-1+4\epsilon})$ . Hence

$$S'' \ll \frac{1}{(\log x)^{2-8\epsilon}} \sum \frac{1}{uv_1}, \quad (25)$$

the latter sum being over all possible values of  $u$  and  $v_1$ .

To estimate the latter, we make the following observations.

- The number  $uv_1$  belongs to  $A$ , satisfies  $y \leq uv_1 \leq x$  and  $\omega(yv_1) \leq k$ .
- Given  $m \in A$  with  $\omega(m) \leq k$ , it can be written as  $m = uv_1$  in at most  $2^k \ll (\log x)^{2\epsilon}$  ways.

It follows that

$$\sum_{uv_1} \frac{1}{m} \ll (\log x)^{2\varepsilon} \sum_{m \in A \cap [y, x]} \frac{1}{m} \ll (\log x)^{6\varepsilon}, \quad (26)$$

the latter sum being  $O((\log x)^{4\varepsilon})$  by Lemma 6.2 with  $b = x$  and  $a = y$ .

Combining (25) and (26), we conclude that

$$S'' \ll \frac{1}{(\log x)^{2-14\varepsilon}}.$$

## 6.6 Proof of Lemma 5.2

Thus, for large  $x$ , the total number of  $n$  such that  $F(n)$  has at least  $6d$  distinct divisors in  $\mathcal{M}_F(x)$  is bounded by

$$x(\log x)^{5\varepsilon \log d} (S' + S'') \ll \frac{x}{(\log x)^{2-5\varepsilon \log d-17\varepsilon}},$$

which proves Lemma 5.2.

## 7 Proof of Theorem 1.6

We are ready now to prove Theorem 1.6. Thus, let  $X$  and  $t$  be as in Theorem 1.6, and, as in Section 3, let  $F(T) \in \mathbb{Z}[T]$  be the primitive separable polynomial whose roots are exactly the finite critical values of  $t$ , with  $d = \deg F$ . We use all notation and conventions from Section 4. In particular, we fix  $\varepsilon$  satisfying  $0 < \varepsilon \leq 1/2$  (which will be specified later) and for sufficiently large  $x$  we consider the set  $\mathcal{M}_F(x)$ .

Recall (see Section 3) that we denote by  $K_n$  the field  $\mathbb{Q}(t^{-1}(n))$ . We call  $m \in \mathcal{M}_F$  primitive for  $n$  if every  $p \mid m$  ramifies in  $K_n$ , but for every  $n' < n$  some  $p \mid m$  does not ramify in  $K_{n'}$ . Clearly, if  $n$  admits a primitive  $m \in \mathcal{M}_F$  then the field  $K_n$  is distinct from  $K_1, \dots, K_{n-1}$ .

Our starting point is Corollary 3.3, which asserts that every  $m \in \mathcal{M}_F$  with the property  $p_{\min}(m) > \omega(m)$  serves as a primitive for some  $n_m \leq m(\omega(m) + 1)$ . If  $m \in \mathcal{M}_F(x)$  then this property is trivially satisfied when  $x$  is large enough; hence every  $m \in \mathcal{M}_F(x)$  serves as primitive for some  $n_m \leq m(k+2)$ , and we have

$$n_m \leq m(k+2) \leq \frac{x}{\log \log x} (\varepsilon \delta \log \log x + 3) \leq x, \quad (27)$$

again provided  $x$  is sufficiently large.

Set

$$\begin{aligned}\mathcal{N}(x) &= \{n_m : m \in \mathcal{M}_F(x)\}, \\ \mathcal{N}'(x) &= \{n \in \mathcal{N}(x) : \text{the fiber } t^{-1}(n) \text{ is } \mathbb{Q}\text{-irreducible}\}.\end{aligned}$$

It follows from (27) that

$$\mathcal{N}'(x) \subset \mathcal{N}(x) \subset [1, x],$$

and Hilbert's Irreducibility Theorem implies that

$$|\mathcal{N}'(x)| \geq |\mathcal{N}(x)| - O(x^{1/2}). \quad (28)$$

The fields

$$K_n \quad (n \in \mathcal{N}(x))$$

are pairwise distinct, and, since for  $n \in \mathcal{N}'(x)$  the field  $K_n$  is the Galois closure of  $\mathbb{Q}(P_n)$ , the fields

$$\mathbb{Q}(P_n) \quad (n \in \mathcal{N}'(x)) \quad (29)$$

are pairwise distinct as well.

Thus, to prove Theorem 1.6, we only have to show that, with suitable choice of  $\varepsilon$ , the lower estimate

$$|\mathcal{N}'(x)| \geq \frac{x}{(\log x)^{1-\eta}} \quad (30)$$

holds for sufficiently large  $x$ . Here  $\eta$  is a positive number depending only on  $d$  (which, through (3), translates into dependence in  $v$  and  $g$ ).

This can be accomplished using the results of Sections 4 and 5. Since every  $p \mid m$  ramifies in  $K_{n_m}$ , we have  $m \mid F(n_m)$  (see Property A in Section 3). Hence Proposition 5.1 applies to our definition of  $n_m$ . Thus, setting  $\varepsilon$  as in (8), Proposition 5.1 implies that, for sufficiently large  $x$ , we have  $|\mathcal{N}(x)| \geq (12d)^{-1} |\mathcal{M}_F(x)|$ . Together with Proposition 4.1 this implies that  $|\mathcal{N}(x)| \geq x(\log x)^{-1+\delta\varepsilon+o(1)}$  as  $x \rightarrow \infty$ , which, combined with (28), implies the same lower estimate for  $|\mathcal{N}'(x)|$ . In particular, for sufficiently large  $x$  we have (30) with  $\eta = \delta\varepsilon/2$ .

In view of (2) and (8) we have  $\eta \geq 10^{-4}(d \log(2d))^{-1}$ . Using (3) we deduce that  $\eta \geq 10^{-6}((g+v) \log(g+v))^{-1}$ .  $\square$

## References

1. YU. BILU, Counting Number Fields in Fibers (with an appendix by JEAN GILLIBERT), submitted; [arXiv:1606.02341\[math.NT\]](#).
2. YU. BILU, F. LUCA, Number Fields in Fibers: the Geometrically Abelian Case with Rational Critical Values, submitted; [arXiv:1606.09164\[math.NT\]](#).
3. P. CORVAJA, U. ZANNIER, On the number of integral points on algebraic curves, *J. reine angew. Math.* **565** (2003), 27–42.
4. H. DAVENPORT, D. LEWIS, A. SCHINZEL, Polynomials of certain special types, *Acta Arith.* **9** (1964), 107–116.
5. J.-M. DE KONINCK, F. LUCA, *Analytic number theory: exploring the anatomy of integers*, Graduate studies in math. 134, AMS, 2012.

6. R. DVORNICICH, U. ZANNIER, Fields containing values of algebraic functions, *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)* **21** (1994), 421–443.
7. R. DVORNICICH, U. ZANNIER, Fields containing values of algebraic functions II (On a conjecture of Schinzel), *Acta Arith.* **72** (1995), 201–210.
8. J.-P. SERRE, *Lectures on the Mordell-Weil Theorem*, 3rd edition, Vieweg & Sohn, Braunschweig, 1997.
9. U. ZANNIER, On the Number of Times a Root of  $f(n, x) = 0$  Generates a Field Containing a Given Number Field, *J. Number Th.* **72** (1998), 1–12.